

VALIDATION OF THE WISCONSIN PERSONALITY DISORDERS INVENTORY-IV WITH THE SCID-II

Tracey L. Smith, PhD, Marjorie H. Klein, PhD,
and Lorna S. Benjamin, PhD

The Wisconsin Personality Disorders Inventory (WISPI-IV; Klein & Benjamin, 1996) is the latest version of a self-report measure of DSM-IV personality disorders (PDs) derived from an interpersonal perspective. When categorical diagnoses derived from the WISPI-IV were compared with independent SCID-II diagnoses, the majority of the kappas were poor ($>.40$). However, all but one of the effect sizes for the differences in WISPI-IV means between groups with and without SCID-II diagnoses were large ($>.80$). When SCID-II and WISPI-IV dimensional scores were considered, the average r between profiles was .61 (median = .58) and correlations between corresponding PD scales (mean diagonal $r = .48$; mean off-diagonal $r = .18$) indicated good convergent and discriminant validity for five of the WISPI-IV scales. These results add to the accumulating evidence suggesting greater reliability and validity of dimensional over categorical scores for PDs. Researchers and clinicians interested in having an efficient method of assessing PDs may consider using a dimensional approach such as the WISPI-IV as an alternative to diagnostic interview.

INTRODUCTION

The purpose of this study was to investigate the validity of the Wisconsin Personality Disorders Inventory (WISPI-IV) using the Structured Clinical Interview for Axis I (SCID-I) as the criterion measure. The WISPI-IV is the most recent version of the original WISPI-III and WISPI-III-R (Klein et al., 1993). Most of the items have remained constant across the three generations of this measure, although the length has been reduced from 302 items for the DSM personality disorder (PD) categories in the original WISPI-III to 204 items in the WISPI-IV. The aim in constructing the WISPI was to create

From the Wisconsin Psychiatric Institute and Clinics, Department of Psychiatry, University of Wisconsin, Madison (T. L. S., M. H. K.), Department of Psychology, University of Utah, Salt Lake City (L. S. B.).

The Mariner S. Eccles Foundation for Political Economy partially supported this research through a fellowship received by the first author during her dissertation study. Special thanks to JuHui Park and Renee Burke for conducting clinical interviews. We gratefully acknowledge the staff and patients at the University Neuropsychiatric Institute and the undergraduate student colleagues who assisted in this project. Thanks to Len Simms for his comments on an earlier draft of this article.

Address correspondence to Tracey L. Smith, Department of Psychiatry, Wisconsin Psychiatric Institute and Clinics, 6001 Research Park Blvd., Madison, WI 53719-1176; E-mail: traceysmith@wisc.edu.

items that presented the DSM criteria for PD in interpersonal terms derived from Benjamin's (1993, 1996) analysis of the DSM items. The basis for this dimensional analysis was the Structural Analysis of Social Behavior model (SASB).

The WISPI scales have demonstrated strong internal consistency. α s for the 11 PD scales of the WISPI-IV ranged .81 to .95 in a mixed sample of student volunteers and psychiatric outpatients (Klein & Benjamin, 1996). These values are similar to those reported for the earlier versions (Barber & Morse, 1994; Klein et al., 1993).

Tests of validity of the earlier versions of the WISPI included tests of content validity from both the DSM and SASB perspectives, comparisons of patients versus nonpatients and patients with and without personality disorders to test discriminant validity, and tests of concurrent validity (Klein et al., 1993). Correlations, corrected for attenuation, averaged .43 for the same personality disorder categories of the WISPI with the Millon Clinical Multiaxial Inventory-I (MCMI-I) and .93 for the WISPI with the Personality Disorder Questionnaire (PDQ). The average of the off-diagonal correlations was .11 for the WISPI-MCMI and .37 for the WISPI-PDQ. In another test of concurrent validity, the average correlation, corrected for attenuation, between the WISPI and clinician ratings on the Personality Assessment Form (PAF) was .51 and .19 for the off-diagonal correlations (Klein et al., 1993). Barber and Morse (1994) compared the WISPI-III-R with the results of SCID-II and Personality Diagnostic Examination (PDE) structured interviews in a mixed outpatient sample and found convergent validity correlations of .44 with the SCID-II and .39 with the PDE. In the Barber and Morse study the highest convergent and discriminant validity was found for avoidant (AVD), obsessive-compulsive (OCD), and borderline (BPD) scales.

All of the above validation analyses were based on dimensional assessments. Barber and Morse (1994) also examined the convergence between a 3-category scoring of the SCID-II (full, subthreshold, absent) and the WISPI-IV and found significant correlations ranging from .27 to .50 for four of the six categories with positive diagnoses (AVD, dependent [DEP], OCD, and passive-aggressive [PAG]).

The SCID-II interview was chosen as the criterion measure for the following reasons: research has shown it is a reliable instrument; it includes both interview and self-report data; it is authoritative in regard to the DSM in part because the SCID-II's authors were central to the development of the DSM; the SCID-II requires shorter administration time; and the SCID-II is widely used.

The internal consistency of the SCID-II PD scales was reported to be high in one study (range .95 to .99; Maffei et al., 1997). Additionally, the interrater reliability of the SCID-II is generally reported to be fair to excellent (kappa range .43 to .98, ICC range .61 to 1.00 for individual PDs; Brooks, Baltazar, McDowell, Munjack, & Bruns, 1991; Maffei et al., 1997; Rennenberg, Chambless, Dowdall, Fauerbach, & Gracely, 1992) and comparable to those obtained by other PD interviews like the Personality Disorders Examination (PDE; Loranger, Susman, Oldham, & Russakoff, 1987) and the Structured Interview for DSM-IV Personality Disorders (SIDP; Stangl, Pfohl, Zimmerman, Bowers, & Corenthal, 1985). Researchers have

found test-retest reliabilities for individual PD diagnoses of .24 to .86 (median .68; Barber & Morse, 1994; Dreesen & Arntz, 1998; First, et al., 1995b), which are also comparable to those attained with other structured PD interviews (First, Spitzer, Gibbon, & Williams, 1995a).

The validity of the SCID-II (both the DSM-III-R and the DSM-IV versions) has been the subject of numerous studies. Lacking an agreed upon "gold standard" in PD assessment, we review the evidence regarding the convergence of the SCID-II with other PD measures. Following the recommendation of Shrout, Spitzer and Fleiss (1997), we consider kappa and ICC values greater than .75 to indicate excellent agreement, values between .74 and .61 as moderate agreement, between .60 and .41 as fair to good, and those below .40 as poor (Shrout, Spitzer, & Fleiss, 1997). Clark, Livesley, & Morey (1997) reviewed 19 studies of the convergence among structured PD interviews and questionnaires and summarized the results by calculating the median kappas across PD categories for the various combinations of measures reviewed. While the convergence of the SCID with interviews and questionnaires was somewhat better than the convergence among the other PD interviews or questionnaires, the levels of agreement were still only in the "moderate" range. To illustrate this we note that only two of the five studies that compared two PD interviews reported median kappas in the "good" range (.46 and .50); both were for comparisons of the SCID-II with the PDE. Convergences between PD interviews and questionnaires were summarized by calculating the median *r*'s or kappas. Only one of the 12 kappas tabled for PD interviews was in the "good" range (.42 for SCID-II with PDQ-R); all others, including three comparisons involving the SCID-II, were "poor," ranging from .38 to .08. Median correlation coefficients between interviews and questionnaires were tabled for 11 studies. Five of the 11 were in the "good" range; one of these was for the SCID-II compared with the WISPI ($r = .46$). None of the median kappas was higher. Three more recent studies compared the SCID-II with questionnaires (PDQ-4+; Fossati et al., 1998; MCMI-II; Kennedy et al., 1995; MCMI-II; Marlowe, Husband, Bonieskie, Kirby, & Platt, 1997). Median correlations were in the "poor" range (.33, .26, and .38, respectively). Therefore, while the SCID-II seems to converge with a number of different PD questionnaires (MCMI-II, PDQ-4+, WISPI) somewhat better than other PD interviews, none of the median *r*'s or kappas reached levels indicative of either moderate or excellent agreement. By summarizing in terms of median kappas or *r*'s we note that convergences may be in acceptable ranges of agreement for some specific PD categories.

There are pros and cons to using self-report questionnaires versus structured clinical interviews for the diagnosis of PD. Assessment of PD by clinical interview requires considerable administrator training, is more time consuming, and thus more costly than PD assessment by self-report instrument. On the other hand, idiosyncratic understanding of items by respondents, inability to query the respondent about state versus trait issues, and the inability to evaluate the evidence that a respondent uses to determine how he or she meets a criterion can hamper assessment of PD by questionnaire. A self-report measure that provides reliable and valid PD diagnostic and symptom information could be useful to both clinicians and researchers. Thus, we examined the reliability and validity of the WISPI-IV questionnaire in relation to the widely used SCID-II interview.

METHOD

PARTICIPANTS

Participants were adult psychiatric inpatients at the University Neuropsychiatric Institute in Salt Lake City, UT who agreed to participate in the research between February 2000 and August 2001. Participants were excluded if they were: currently psychotic or receiving electroconvulsive therapy; had organic brain damage; were identified as mentally handicapped; did not speak English as a native language; or were hospitalized primarily for the treatment of alcohol or drug abuse.

Of the patients approached, 125 agreed to participate. Fifteen of these individuals completed the initial screening but left the hospital either before they could complete the full assessment or declined to complete the procedure. Of the remaining 110 patients, 75 completed both the WISPI and the SCID-II.¹ Table 1 shows the demographic information and the SCID Axis I and II diagnoses for these 75 participants. Participants' mean age was 35.29 years ($SD = 11.39$). Participants ranged from the very lowest to the highest scores possible (12 - 66) on the Four Factor Index of Social Status (Hollingshead, 1975). The mean was 41 ($SD = 12.54$) commensurate with medium to small size business owners, minor professionals, and technical workers. Over 95% of the participants were white, which is consistent with the ethnic makeup of the region and the patients attending the hospital where data were collected.

PROCEDURE

The first author and a graduate student colleague reviewed the charts of patients on the open inpatient unit for the presence of behavior or symptoms consistent with DSM-IV personality pathology and absence of exclusion criteria. Researchers explained the purpose of the research, the data collection procedure, and the compensation for participation. Patients who indicated initial interest were given consent forms to read and sign. Researchers also offered to send a copy of the participant's computer-generated diagnostic assessments to their psychiatrist or licensed therapist. There were three computer-administered tests, a demographics questionnaire, the SCID-I Screen Patient Questionnaire (First, Gibbon, Spitzer, & Williams, 1997), and the SCID-II Patient Questionnaire (SCID-II PQ; First, Gibbon, Spitzer, Williams, & Benjamin, 1996). After screening, patients were interviewed using the Structured Clinical Interviews for Axis I and II (SCID-I, Patient Edition, version 2.0; First, Spitzer, Gibbon, and Williams, 1997; SCID-II; First, Gibbon, Spitzer, Williams, & Benjamin, 1997). Following standard SCID-II

1. In the total sample, no differences were found between completers and noncompleters on any of the demographic variables [age, $t(121) = -.66$; ns, SES, $t(114) = -1.07$, ns; sex, $\chi^2(1, N = 124) = .94$, ns; marital status, $\chi^2(2, N = 119) = 1.59$, ns; or education, $\chi^2(3, N = 120) = 4.41$, ns.] Nor did people who completed all or most of the self-report measures differ from those who did not complete them in age, $t(121) = -.61$; ns, socioeconomic status, $t(114) = -.31$, ns; sex, $\chi^2(1, N = 124) = .00$, ns; marital status, $\chi^2(2, N = 119) = 4.49$, ns; or education, $\chi^2(3, N = 120) = 2.35$, ns.

TABLE 1. Sample Characteristics

	Number (%)
Sex	
Female	56 (74.7)
Male	19 (25.3)
Marital Status	
Single	23 (30.7)
Partner	35 (46.7)
Sep., Div., Widow.	16 (21.3)
Missing	1 (1.3)
SES	
Hollingshead 1 & 2	12 (16.0)
Hollingshead 3	18 (24.0)
Hollingshead 4	33 (44.0)
Hollingshead 5	9 (12.0)
Missing	3 (4.0)
Education	
HS or less	22 (29.3)
College degree	8 (10.7)
1 – 3 college	34 (45.3)
5 + college	11 (14.7)
DSM–IV Axis I Diagnoses†	
Major Depressive	56 (74.7)
Bipolar II	10 (13.3)
Panic Disorder	27 (36.0)
OCD	9 (12.0)
Dysthymia	27 (36.0)
Bipolar I	7 (9.3)
PTSD	18 (24.0)
Bulimia	7 (9.3)
Generalized Anxiety	16 (21.3)
Binge Eating	6 (8.0)
Social Phobia	13 (17.3)
Pain Disorder	6 (8.0)
DSM–IV Axis II Diagnoses	
Avoidant (AVD)	28 (37.3)
Borderline (BPD)	43 (55.8)
Obsessive–Com. (OCD)	35 (46.7)
Passive–Aggressive (PAG)	22 (29.3)
Paranoid (PAR)	19 (25.3)
Dependent (DEP)	9 (12.0)
Schizoid (SZD)	1 (1.3)
Schizotypal (SZT)	3 (4.0)
Narcissistic (NAR)	3 (4.0)
Histrionic (HST)	0
Antisocial (ASP)	0

Note. Sep., Div., Widow. – Separated, divorced, or widowed. †DSM–IV diagnoses with prevalence > 5 in this sample, other diagnoses not listed. N = 75.

protocol, interviewers only queried items patients endorsed on the SCID-II screen. Interviewers queried unendorsed items when (1) a participant was one criterion away from meeting a PD diagnosis, or (2) the participant demonstrated or disclosed information that was consistent with an unendorsed item. Research has shown that following this procedure does not result in appreciable rates of false negatives despite initial concerns (Jacobsberg, Perry, & Frances, 1995). The primary investigator reviewed each SCID-I and II interview for coding errors or missed items within two days of administration so that errors could be resolved with the patient or interviewer as needed.

The primary investigator completed 57% percent of the interviews, 37% were completed by a second graduate student, and 6% by a third graduate student. After establishing initial interrater reliability ($\kappa = .92$; for presence versus absence of a disorder on both the SCID-I and SCID-II), about every tenth interview of each interviewer was videotaped and coded for reliability (based on patient consent). Average κ for Axis I disorders across all videotapes was .99 and for Axis II disorders was .96. The ICC value for the interrater reliability of the dimensional SCID-II scores was .87.

WISPI-IV

Participants were also given the WISPI-IV in paper-and-pencil form (Klein & Benjamin, 1996). This self-report questionnaire provides both categorical diagnoses and dimensional scores for 11 PD categories. There are 204 items that are rated on a 10-point scale that ranges from 0 (never or not at all true of you) to 10 (always or extremely true of you). Ten of the items are from the Marlowe-Crowne Scale for social desirability (Greenwald & Satow, 1970). Each item is written from the point of view of the respondent who is asked to rate their "usual self" during the past 5 years or more. Two of the scores created by the WISPI-IV scoring program were used in the data analyses: (a) mean scores (the means of the ratings for the items on each scale) and (b) z-scores (computed using the normal sample data from Klein et al., 1993). For the concurrent validity analyses using PD categories, we used two methods of assigning a categorical PD diagnosis from the WISPI-IV. In the first method, patients were assigned a diagnosis if they rated at least one item at the level of 6 or higher for the minimum number of DSM-IV criteria needed for each PD category. The second method assigned a diagnosis if a patient had a z-score of 1.96 or greater on a PD scale (i. e., their score was significantly greater than the normative sample at $p < .05$).

RESULTS

CHARACTERISTICS OF THE DATA SET

All variables were examined for statistical anomalies and to ensure that statistical assumptions were met. Some participants returned WISPI-IV questionnaires with missing data on a few items. The amount of missing data on the WISPI-IV was less than 2% across all items. Mean values across all par-

TABLE 2. Comorbidity between PD on the SCID-II

MET	AVD	DEP	OCD	PAG	PAR	SZT	SZD	HST	NAR	BPD	ASP	No PD
AVD	—											
DEP	5	—										
OCD	6	4	—									
PAG	10	3	15	—								
PAR	8	3	11	13	—							
SZT	2	1	0	2	2	—						
SZD	1	0	0	0	0	0	—					
HST	0	0	0	0	0	0	0	—				
NAR	0	0	2	2	2	0	0	0	—			
BPD	16	5	21	16	17	3	0	0	3	—		
ASP	0	0	0	0	0	0	0	0	0	0	—	
No PD	—	—	—	—	—	—	—	—	—	—	—	5

Note. BPD = Borderline; AVD = Avoidant; OCD = Obsessive-Compulsive; DEP = Dependent; PAG = Passive-Aggressive; PAR = Paranoid; SZT = Schizotypal; SZD = Schizoid; HST = Histrionic; NAR = Narcissistic; ASP = Antisocial. *N* = 75.

ticipants were used to replace missing values. For the total sample, the mean values for the WISPI-IV PD scales were within a standard deviation of those reported for a patient sample that included inpatients and outpatients (Klein et al., 1993).

DIAGNOSES

The SCID-II can provide both categorical diagnoses and dimensional scores for each PD scale. Each SCID-II item is rated 1 (absent), 2 (subthreshold), or 3 (threshold). An individual receives a diagnosis when he or she meets or exceeds the minimum number of criteria at threshold for a given PD category. To create a dimensional score from the SCID-II data, we calculated the percentage of weighted criterion scores for each PD scale divided by the total possible. For example, if a patient met 4 criteria at threshold level (for a weight of 3) on the AVD scale, which has a possible total score of 21, their score would be 57%. Table 2 shows the comorbidity between PD groups. The mean number of PDs per participant was 2.19 (*SD* = 1.44; range 0 - 6) and the mean number of criteria met was 21.32 (*SD* = 9.27; range 5 - 44). This is lower than the generally reported average of four PDs per patient (e. g., Skodol, Rosnick, Kellman, Oldham, & Hyler, 1991; Oldham et al., 1992).

RELIABILITY OF THE WISPI-IV AND THE SCID-II

Table 3 presents the means, standard deviations, and Cronbach's α s for the dimensional SCID-II and the WISPI-IV PD scales. Internal consistency for the WISPI-IV was high, with α s ranging from .74 (Antisocial PD; ASP) to .91 (AVD). SCID-II α s were lower and ranged from .38 (ASP) to .77 (AVD). The lack of diagnosed cases of ASP on the SCID-II in this sample likely contributed to the lower reliability coefficients for this scale. Very few patients met

TABLE 3. Means, Standard Deviations, & Cronbach's Alpha for the Percentage of SCID-II Weighted Sum Total, & Mean and Percentage WISPI Scale Scores

	SCID-II		WISPI-IV		
	Mean % (SD)	Alpha	Mean (SD)	Mean % (SD)	Alpha
BPD	.50 (.30)	.74	3.97 (1.64)	.40 (.16)	.87
AVD	.42 (.30)	.77	5.63 (1.91)	.52 (.21)	.91
OCD	.46 (.26)	.56	3.83 (1.26)	.37 (.12)	.80
DEP	.18 (.20)	.59	3.85 (1.85)	.37 (.18)	.91
PAG	.33 (.27)	.58	2.96 (1.00)	.30 (.12)	.84
PAR	.29 (.25)	.66	3.78 (1.58)	.37 (.17)	.88
SZT	.15 (.17)	.63	2.42 (1.08)	.24 (.12)	.82
SZD	.11 (.14)	.52	3.19 (1.28)	.31 (.12)	.79
HST	.08 (.11)	.30	3.09 (1.50)	.34 (.17)	.89
NAR	.16 (.20)	.68	2.88 (1.27)	.29 (.12)	.84
ASP	.08 (.16)	.38	1.58 (0.58)	.16 (.01)	.74

Note. BPD = Borderline; AVD = Avoidant; OCD = Obsessive-Compulsive; DEP = Dependent; PAG = Passive-Aggressive; PAR = Paranoid; SZT = Schizotypal; SZD = Schizoid; HST = Histrionic; NAR = Narcissistic; ASP = Antisocial. *N* = 75.

criteria C on this scale (evidence of Conduct Disorder before the age of 15) so interviewers, who followed the skip out instructions, did not query the remaining items for this scale.

WISPI-IV INTERSCALE CORRELATIONS

In the DSM-III-R version of the WISPI, interscale correlations were quite high with an overall average of .62 and a range of .48 (SZD, Schizoid) to .69 (PAR, Paranoid; Klein et al., 1993). To compute average correlations, we converted *r* values to *z* scores using Fisher's *r* to *z* conversion, averaged the *z* scores, and then converted the average *z* back to an *r*. In this sample, the overall average interscale correlation for the WISPI-IV was .46 with a range of .33 (ASP) to .56 (Narcissistic, NAR). While the degree of interscale correlation has been reduced in this latest version of the WISPI, the interscale correlations are still relatively high.

WISPI-IV & SCID-II CORRESPONDENCE ON CATEGORICAL DIAGNOSES

Table 4 shows the agreement between categorical diagnoses on the two instruments for each of the six PDs in the sample that had more than five cases as determined by the SCID-II interview. The top of Table 4 presents WISPI-IV-SCID-II convergences when the WIPSI-IV's categorical DSM-IV scoring procedures were used. The correct classification rate ranged from .55 (OCD) to .78 (PAR) with a mean of .67. Kappa values were poor to moderate and ranged from .08 to .48 (*M* = .26). The positive predictive power (PPP) ranged from .22 (DEP) to .76 (BPD) whereas the negative predictive power (NPP) ranged from .59 (BPD) to .94 (DEP).

TABLE 4. Interinstrument Agreement between PD with More than Five Cases on the SCID-II Using Two Methods for Establishing PD Diagnosis: WISPI-IV DSM-IV Diagnostic Variable & z Values

	SCID-II	WISPI						Class.
	Prev.	Prev.	κ	Sensit.	Specific.	PPP	NPP	Rate
WISPI-IV DSM-IV Diagnostic Variable†								
AVD	28	46	.34	.86	.53	.52	.86	.65
DEP	9	27	.19	.67	.68	.22	.94	.68
OCD	35	29	.08	.43	.65	.52	.57	.55
PAG	22	17	.14	.32	.81	.41	.74	.67
PAR	20	27	.48	.76	.79	.26	.90	.78
BPD	43	34	.38	.60	.75	.76	.59	.68
WISPI-IV z Values‡								
AVD	28	27	.63	.75	.87	.78	.85	.83
DEP	9	18	.25	.56	.80	.28	.93	.77
OCD	35	8	.13	.17	.95	.75	.57	.59
PAG	22	8	.13	.18	.92	.50	.73	.71
PAR	20	13	.29	.37	.89	.54	.81	.73
BPD	43	21	.30	.42	.91	.86	.54	.63

Note. Prev. = prevalence. AVD = Avoidant; DEP = Dependent; OCD = Obsessive-Compulsive; PAG = Passive-Aggressive; PAR = Paranoid; BPD = Borderline. Sensit. = sensitivity, percentage of patients correctly diagnosed by the WISPI-IV; Specific = specificity, percentage of patients correctly not diagnosed by the WISPI-IV. PPP = Positive Predictive Power, probability that patient has a PD when positive for the diagnosis on the WISPI-IV. NPP = Negative Predictive Power, probability that patient does not have a PD when negative for the diagnosis on the WISPI-IV. Class Rate = classification rate, percentage of total sample correctly identified by the WISPI-IV. †DSM-IV Diagnostic variable = met at least one item for the minimum number of criteria needed to attain a categorical diagnosis. ‡WISPI-IV z values = z values created from normals' data in the validation sample (Klein et al., 1993). $N = 75$.

The bottom of Table 4 shows the interinstrument agreement using WISPI-IV cut-off z scores of 1.96 to establish diagnoses. This increased the correct classification rate to a mean of .71, kappa values increased slightly (range .13 - .63; $M = .29$). In general, the convergent validity of categorical diagnoses between the two instruments ranged from poor to good, as was generally the case in the cross-method assessments of PD reviewed (Clark, Livesley, & Morey, 1997; Perry, 1992).

Table 5 shows the Cohen effect sizes (d) when the WISPI-IV mean scale scores are compared between the participants who met criteria for a particular PD on the SCID-II and those who did not. This statistic describes the distance between the means of the disordered and nondisordered groups in pooled standard deviation units (Cohen, 1988; Hsu, 2002). Table 5 also presents two measures of the percentage of overlap and nonoverlap between the two distributions of WISPI-IV mean scores. Here, the WISPI-IV distinguished between those who were diagnosed with a particular PD on the SCID-II and those who were not, as demonstrated by the large effect sizes ($d .80$; Cohen, 1988) for five of the six PDs that had more than five diagnosed cases on the SCID-II.

TABLE 5. Cohen Effect Sizes and Measures of Overlap and Nonoverlap Based on WISPI-IV Mean Scale Scores for PDs with More than Five Diagnosed Cases on the SCID-II

	M Dx on SCID-II	WISPI z	M Not Dx on SCID-II	d	% Nonoverlap	% of PD Above M of Non PD
BPD	4.70	1.68	2.93	1.10	58.9	86.4
AVD	6.43	2.18	4.49	.90	51.6	81.6
OCD	3.98	.55	3.33	.52	33.0	69.1
DEP	4.99	1.93	2.71	.82	47.4	78.8
PAG	3.66	.94	2.71	.81	47.4	78.8
PAR	4.86	1.54	3.15	1.05	58.9	86.1

Note. M Dx on SCID-II = Mean WISPI-IV scale score for group diagnosed on the SCID-II. WISPI-IV z value = Mean WISPI-IV z score for group diagnosed on the SCID-II. *d* = Cohen's Effect Size. % Nonoverlap = percentage of nonoverlap between the SCID-II diagnosed PD and not diagnosed PD groups. % of PD Above = percentage of patients with a particular SCID-II diagnosed PD scoring above the mean of those patients not diagnosed on the SCID with that particular PD.

CORRESPONDENCE BETWEEN THE SCID-II DIMENSIONS AND THE WISPI-IV

To examine the relationship between SCID-II and WISPI-IV dimensional scores, the SCID-II dimensional (percentage) scores that we calculated were correlated with the means of the WISPI-IV PD scales. As Table 6 shows, 10 out of 11 dimensional SCID-II PD scale scores had their highest correlation (indicated in underlined type, follow rows) with the corresponding WISPI-IV scales (except Histrionic, HST). This indicates that the two DSM-IV measures converged to a reasonable degree (mean diagonal $r = .48$; mean off-diagonal $r = .18$). Five out of 11 WISPI-IV scores had their highest correlation (indicated in bold type, follow columns) with the corresponding SCID-II scales. Good convergent and discriminant validity (discriminating from the other PD scales; Trull, 1993) was demonstrated on five PD scales (BPD, AVD, DEP, PAR, and schizotypal [SZT]). For these PDs, the highest correlation on both the SCID-II and the WISPI-IV was with the corresponding scale on the other measure. Unlike the study by Barber and Morse (1994), the SCID-II PAR scale, rather than the DEP scale, was significantly correlated with many of the WISPI-IV PD scales. This difference may be attributable to differences in our participant populations. Barber and Morse recruited outpatient participants with specific Axis I and II disorders² whereas the participants in our sample were psychiatric inpatients on the unlocked ward. The results of this study are consistent with the convergence between the SCID-II and WISPI-IV shown in the Barber and Morse study. Their mean correlation between corresponding scales (using earlier versions of the SCID-II and the WISPI) was .44 as compared to our .48.

2. Patients were selected if they had diagnoses of chronic depression, general anxiety disorder, obsessive-compulsive personality, or avoidant personality disorder.

TABLE 6. Correlations Between SCID-II Dimensional Scores & Mean WISPI Scale Scores

	WISPI-IV										
	BPD	AVD	OCD	DEP	PAG	PAR	SZT	SZD	HST	NAR	ASP
SCID-II											
BPD	.60**	.16	.13	.29*	.35**	.33**	.13	.11	.29*	.42**	.34**
AVD	.21	.60**	.05	.34**	.18	.10	.11	.20	.16	.02	.13
OCD	.01	.07	.44**	.09	.14	.22	.07	.14	.03	.24*	.01
DEP	.37**	.26*	.16	.53**	.17	.11	.29*	.12	.20	.12	.10
PAG	.25*	.10	.32	.12	.38**	.26*	.01	.17	.17	.34**	.24*
PAR	.58**	.31**	.47**	.16	.43*	.60**	.25*	.43**	.25*	.56**	.40**
SZT	.26*	.29*	.03	.16	.13	.21	.32**	.28*	.11	.20	.13
SZD	.04	.08	.07	.11	.14	.02	.06	.36**	-.24*	.10	.03
HST	.45**	.02	.13	.30**	.27*	.25*	.08	.03	.43**	.32**	.33**
NAR	.46**	.75	.29*	.20	.44**	.29*	.11	.06	.49**	.54**	.44**
ASP	.14	.21	.17	.19	.07	.19	.01	.12	.02	.19	.40**

Note. BPD = Borderline; AVD = Avoidant; OCD = Obsessive-Compulsive; DEP = Dependent; PAG = Passive-Aggressive; PAR = Paranoid; SZT = Schizotypal; SZD = Schizoid; HST = Histrionic; NAR = Narcissistic; ASP = Antisocial. *N* = 75. **p* < .05; ***p* < .01.

PROFILE CORRELATIONS

Another way of examining convergence between the two PD measures is to examine the correspondence between the profiles of the 11 PD scales on each measure for each participant. This within-subject procedure considers all of the PD dimensions at once and derives an index of overall congruence (see Figure 1). To make the WISPI-IV scores metrically commensurate with the SCID percentage scores, we calculated WISPI percentages from the sum of all item endorsements within a scale, divided by the total score possible for that scale. After reversing the data matrix (participants become the columns and the 11 PDs the rows), each participant's profile of percentage dimensional scores on the SCID-II was correlated with his or her corresponding WISPI-IV profile of percentage scores. The resulting statistic (a within-subjects Pearson product moment correlation) represents the congruence of the two measures across all PDs per patient. The average *r* (after *r* to *z* conversion as described previously) between the congruence scores across all participants was .61 with a range of -.01 to .93 (median = .58). This is consistent with an earlier study of the WISPI and SCID-II, which found a mean profile correlation of .53 (Barber & Morse, 1994).

DISCUSSION

In many respects, the results of this study replicate and improve upon previous research on the WISPI (Barber & Morse, 1994; Klein et al., 1993). Consistent with this past research, this study demonstrates that the theory-based WISPI-IV has high internal consistency and good convergent (with other measures of PD) and divergent (between PD scales) validity with the SCID-II interview (Barber & Morse, 1994) and other PD assessment measures (Klein et al.,

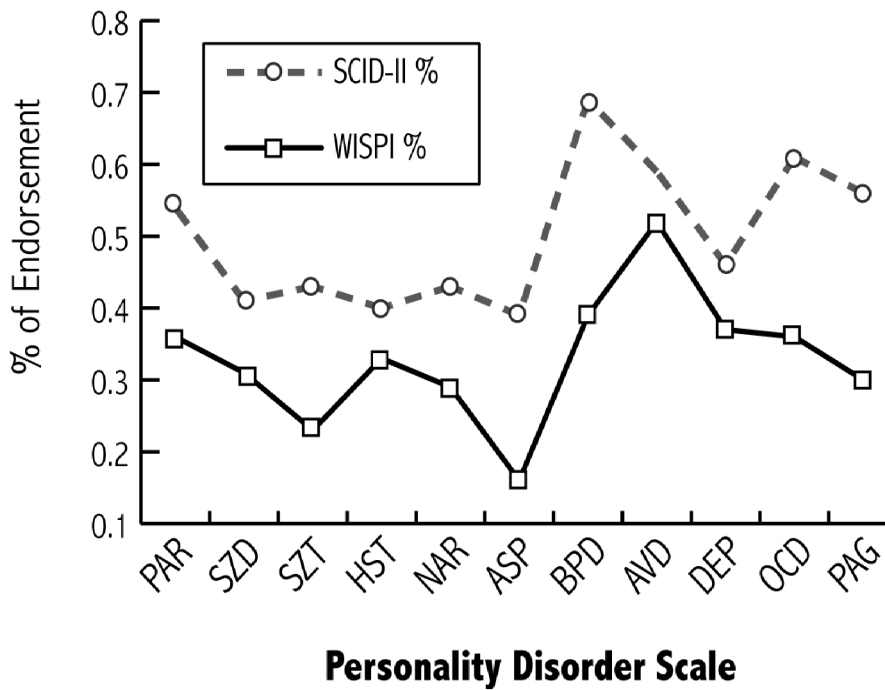


FIGURE 1. Profiles of mean percentage of endorsement on the WISPI-IV and SCID-II PD scales.

1993). This research has shown that test-retest reliability on the WISPI is adequate (.70; as suggested by Nunnally & Bernstein, 1994) even over a 3- to 4-month period (Barber & Morse, 1994). Temporal stability on a PD measure is important given that PDs are by definition “an enduring pattern of inner experience or behavior that ... is stable over time” (APA, 1994, p. 630). Interscale correlations between WISPI-IV scales are lower in this latest version of the measure suggesting that the various PD scales are capturing unique aspects of PD pathology. Klein and colleagues (1993) reported an average interscale correlation of .62 (range .48 - .69) for the WISPI-III compared to the current results of .46 (range of .33 - .56) for the WISPI-IV.

Before discussing the convergence between the WISPI-IV and the SCID-II, some limitations of this study should be noted. Our sample consisted of psychiatric inpatients in one particular hospital. This may have had an effect on the types and levels of personality disorders in the sample. Additionally, patients admitted primarily for substance abuse were excluded, which likely affected the prevalence of some PDs such as ASP. It should be noted however, that the inpatients in the study were not free from current or past substance abuse. Examination of the results of the SCID-I screening questionnaire revealed that 21% of the inpatients endorsed all three screening questions related to alcohol abuse, 23% stated they had tried street drugs, and 32% said they had been “hooked on” a prescribed drug or had

taken more of it than was prescribed. The sample was exclusively Caucasian. Together these sample characteristics limit the ability to generalize these findings to other populations. Additionally, this sample only had 6 PDs with more than 5 SCID-II diagnosed cases out of the 11 PDs represented on the WISPI-IV, which limited our ability to examine categorical diagnostic efficiency to just those 6 PDs. The lack of established prevalence rates for the various PDs at this hospital prevented us from using some diagnostic statistics recently recommended by Hsu (2000). Given the characteristics of the particular unit where data was collected (an unlocked ward), we decided it was untenable to estimate inpatient prevalence rates from the research literature. Perhaps the most important limitation, as with most studies of PD, is the lack of an independent or consensual gold (or even LEAD) standard diagnosis to use as the criterion.

Our results add to the wealth of evidence that dimensional assessment of PDs results in higher validity and reliability data than does categorical assessment (Barber & Morse, 1994; Klein, 1993; Widiger, 1993). We note that the low kappas obtained for the comparisons between SCID-II and WISPI-IV categorical diagnoses are quite consistent with prior research (e.g., Clark, Livesley, and Morey, 1997). In a recent overview of diagnostic validity statistics, Hsu (2002) noted that kappa (and other related statistics such as PPP and NPP) has been "criticized because of its sensitivity to base rates" (p. 413). While recognizing the clinical appeal and utility of kappas, he suggests other promising methods (including effect size), which use more robust dimensional data to enhance the diagnostic process.

Examination of the correlations between SCID-II dimensional scores and the WISPI-IV scales demonstrated good convergence (mean r between corresponding PD scales = .48). Discriminant (from other PD scales) as well as convergent validity was demonstrated for BPD, AVD, DEP, PAR, and SZT scales suggesting that these scales can be more easily distinguished from other PDs. Profile analysis provided the most compelling evidence for convergence (mean r = .61) demonstrating that individuals show a similar pattern of responses across the 11 PD scales on the 2 measures. The results of this profile analysis were marginally stronger than those previously reported between these two measures (r = .61 vs. .53 in Barber & Morse, 1994). This correspondence suggests that researchers and clinicians interested in having an inexpensive and time saving method of assessing the relative degree to which individuals have characteristics consistent with various PDs may want to consider using the WISPI-IV as an alternative to the SCID-II.

REFERENCES

- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders*. (4th ed.). Washington, DC: Author.
- Barber, J. P. & Morse, J. Q. (1994). Validation of the Wisconsin Personality Disorders Inventory with the SCID-II and PDE. *Journal of Personality Disorders*, 8, 307-319.
- Benjamin, L. S. (1993). *Interpersonal diagnosis and treatment of personality disorders*. New York: Guilford Press.
- Benjamin, L. S. (1996). *Interpersonal diagnosis and treatment of personality disorders* (2nd ed.). New York: Guilford.
- Brooks, R. B., Baltazar, P. L., McDowell, D. E., Munjack, D. J., & Bruns, J. R. (1991). *Personality disorders*

- co-occurring with panic disorder with agoraphobia. *Journal of Personality Disorders*, 5, 328-336.
- Clark, L. A., Livesley, W. J., & Morey, L. (1997). Personality disorder assessment: The challenge of construct validity. *Journal of Personality Disorders*, 11, 205-231.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dreessen, L., & Arntz, A. (1998). Short-interval test-retest interrater reliability of the Structured Clinical Interview for DSM-III-R personality disorders (SCID-II) in outpatients. *Journal of Personality Disorders*, 12, 138-148.
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1996). *Structured Clinical Interview for the DSM-IV Axis I Personality Disorders (SCID-I/P)*. New York: New York State Psychiatric Institute.
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B., & Benjamin, L. S. (1997). *User's guide for the Structured Clinical Interview for the DSM-IV Personality Disorders*. Washington, DC: American Psychiatric Press.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995a). The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II). Part I: Description. *Journal of Personality Disorders*, 9, 83-91.
- First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W., Davies, M., Borus, J., et al. (1995b). The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II): Part II: Multi-site test-retest reliability study. *Journal of Personality Disorders*, 9, 92-104.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1997). *Structured Clinical Interview for the DSM-IV Axis I Disorders - Patient Edition (SCID-I/P, Version 2.0, 4/97 revision)*. New York: New York State Psychiatric Institute.
- Fossati, A., Maffei, C., Bagnato, M., Donati, D., Donini, M., Fiorilli, M., Novella, L., & Ansoldi, M. (1998). Criterion validity of the Personality Diagnostic Questionnaire-4+ (PDQ-4+) in a mixed psychiatric sample. *Journal of Personality Disorders*, 12, 172-178.
- Greenwald, H. J., & Satow, Y. (1970). A short social desirability scale. *Psychological Reports*, 27, 131-135.
- Hollingshead, A. B. (1975). *Four Factor Index of Social Status*. Unpublished manuscript, Yale University, Department of Sociology, New Haven, CT.
- Hsu, L. M. (2002). Diagnostic Validity Statistics and the MCMI-III. *Psychological Assessment*, 14, 410-422.
- Jacobsberg, L., Perry, S., & Frances, A. (1995). Diagnostic agreement between the SCID-II Screening Questionnaire and the Personality Disorder Examination. *Journal of Personality Assessment*, 65, 428-433.
- Kennedy, S. H., Katz, R., Rockert, W., Mendolwitz, S., Raleveski, E., & Clewes, J. (1995). Assessment of personality disorders in anorexia nervosa and bulimia nervosa. A comparison of self-report and structured interview methods. *Journal of Nervous and Mental Disease*, 183, 358-364.
- Klein, M. H. (1993). Issues in the assessment of personality disorders. *Journal of Personality Disorders*, Supplement, 18-33.
- Klein, M. H. & Benjamin, L. S. (1996). *The Wisconsin Personality Disorders Inventory-IV*. Madison, WI: University of Wisconsin, unpublished test. Available from Dr. M. H. Klein, Department of Psychiatry, Wisconsin Psychiatric Institute and Clinic, 6001 Research Park Blvd., Madison, WI 53719-1179.
- Klein, M. H., Benjamin, L. S., Rosenfeld, R., Treece, C., Husted, J., & Greist, J. H. (1993). The Wisconsin Personality Disorders Inventory: I: Development, reliability, and validity. *Journal of Personality Disorders*, Supplement, 18-33.
- Loranger, A. W., Susman, V. L., Oldham, J. M., & Russakoff, L. M. (1987). Personality Disorder Examination: A preliminary report. *Journal of Personality Disorders*, 1, 1-13.
- Maffei, C., Fossati, A., Agostoni, I., Barraco, A., Bagnato, M., Donati, D. et al. (1997). Interrater reliability and internal consistency of the structured clinical interview for DSM-IV Axis I personality disorders (SCID-II), version 2.0. *Journal of Personality Disorders*, 11, 279-284.
- Marlowe, D. B., Husband, S. D., Bonieskie, L. M., Kirby, K. C., & Platt, J. J. (1997). Structured interview versus self-report test advantages for the assessment of personality pathology in cocaine dependence.

- Journal of Personality Disorders*, 11, 177-190.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill, Inc.
- Oldham, J. M., Skodol, A. E., Kellman, H. D., Hyler, S. E., Rosnick, L., & Davies, M. (1992). Diagnosis of DSM-III-R personality disorders by two structured interviews: Patterns of comorbidity. *American Journal of Psychiatry*, 149, 213-220.
- Perry, J. C. (1992). Problems and considerations in the valid assessment of personality disorders. *American Journal of Psychiatry*, 149, 1645-1653.
- Rennenberg, B., Chambless, D. L., Dowdall, D. J., Fauerbach, J. A., & Gracely, E. J. (1992). The Structured Clinical Interview for DSM-III-R, Axis I and the Millon Clinical Multiaxial Inventory: A concurrent validity study of personality disorders among anxious outpatients. *Journal of Personality Disorders*, 6, 117-124.
- Shrout, P. E., Spitzer, R. L., & Fleiss, J. L. (1997). Quantification in psychiatric evaluation revisited. *Archives of General Psychiatry*, 44, 172-177.
- Stangl, D., Pfohl, B., Zimmerman, M., Bowers, W., & Corenthal, C. (1985). A structured interview for the DSM-III personality disorders: A preliminary report. *Archives of General Psychiatry*, 42, 591-596.
- Skodol, A. E., Rosnick, L., Kellman, H. G., Oldham, J. M., & Hyler, S. E. (1991). Development of a procedure for validating structured assessments of Axis I. In J. Oldham (Ed.), *Personality disorders: New perspectives on diagnostic validity* (pp. 41-70). Washington, DC: American Psychiatric Press.
- Trull, T. J. (1993). Temporal stability and validity of two personality disorder inventories. *Psychological Assessment*, 5, 11-18.
- Widiger, T. A. (1993). The DSM-III-R categorical personality disorder diagnoses: A critique and an alternative. *Psychological Inquiry*, 4, 75-90.